

Develop Foreign Language Corpora



Project Title	Develop Foreign Language Corpora
Project Summary	Develop a corpus of texts in foreign languages, using sources identified in coordination with the School of Language Studies at the Foreign Service Institute, the training arm of the State Department.
Country	United States
Agency	Department of State
DoS Office	FSI/SLS/CSD
Post	N/A
Section	N/A
Number of Interns	50

Project Description

Develop a corpus of texts in foreign languages, using sources identified in coordination with the School of Language Studies. The texts would be important for the kind of work done by Foreign Service Officers and reflect the five professional cones used by the department. The goal of the project is to produce lists of high frequency words that support curriculum development work in the Language School. The intern will develop corpora (database) of foreign language texts containing at least 2,500,000 tokens (words), and conduct corpus analysis using AntWord Profiler and AntConc, two freely available software packages. The curriculum unit will provide guidance and coaching during the project.

Desired Skills Interests

Additional Information

You need to be able to read in at least one of the language requirements listed.

You will need to use AntWordProfiler and AntConc for this project, using your own computer. Both are available free from <http://www.laurenceanthony.net/>

Language Requirements

Language	Speaking Proficiency	Reading Proficiency	Importance
Armenian	No practical proficiency	Limited working proficiency	Nice to Have
Arabic	No practical proficiency	Limited working proficiency	Nice to Have
Bosnian	No practical proficiency	Limited working proficiency	Nice to Have
Amharic	No practical proficiency	Limited working proficiency	Nice to Have
Bulgarian	No practical proficiency	Limited working proficiency	Nice to Have
Czech	No practical proficiency	Minimum professional proficiency	Nice to Have
Estonian	No practical proficiency	Limited working proficiency	Nice to Have
Georgian	No practical proficiency	Limited working proficiency	Nice to Have
Greek	No practical proficiency	Limited working proficiency	Nice to Have
German	No practical proficiency	Limited working proficiency	Nice to Have
Gujarati	No practical proficiency	Limited working proficiency	Nice to Have
Haitian Creole	No practical proficiency	Limited working proficiency	Nice to Have
Hungarian	No practical proficiency	Limited working proficiency	Nice to Have
Hungarian	No practical proficiency	Limited working proficiency	Nice to Have
Indonesian	No practical proficiency	Limited working proficiency	Nice to Have
Italian	No practical proficiency	Limited working proficiency	Nice to Have
Japanese	No practical proficiency	Limited working proficiency	Nice to Have
Kazakh	No practical proficiency	Limited working proficiency	Nice to Have
Kirghiz	No practical proficiency	Limited working proficiency	Nice to Have
Kurdish	No practical proficiency	Limited working proficiency	Nice to Have
Lao	No practical proficiency	Limited working proficiency	Nice to Have
Lithuanian	No practical proficiency	Limited working proficiency	Nice to Have
Macedonian	No practical proficiency	Limited working proficiency	Nice to Have
Malay	No practical proficiency	Limited working proficiency	Nice to Have
Mongolian	No practical proficiency	Limited working proficiency	Nice to Have
Nepali/Nepalese	No practical proficiency	Limited working proficiency	Nice to Have
Pashto	No practical proficiency	Limited working proficiency	Nice to Have
Persian-Dari	No practical proficiency	Limited working proficiency	Nice to Have
Persian-Tajiki	No practical proficiency	Limited working proficiency	Nice to Have
Polish	No practical proficiency	Limited working proficiency	Nice to Have
Portuguese	No practical proficiency	Limited working proficiency	Nice to Have
Romanian	No practical proficiency	Limited working proficiency	Nice to Have

Serbo-Croatian	No practical proficiency	Limited working proficiency	Nice to Have
Slovak	No practical proficiency	Limited working proficiency	Nice to Have
Singhalese	No practical proficiency	Limited working proficiency	Nice to Have
Somali	No practical proficiency	Limited working proficiency	Nice to Have
Swahili/Kiswahili	No practical proficiency	Limited working proficiency	Nice to Have
Swedish	No practical proficiency	Limited working proficiency	Nice to Have
Telugu	No practical proficiency	Limited working proficiency	Nice to Have
Thai	No practical proficiency	Limited working proficiency	Nice to Have
Tibetan	No practical proficiency	Limited working proficiency	Nice to Have
Turkish	No practical proficiency	Limited working proficiency	Nice to Have
Ukrainian	No practical proficiency	Limited working proficiency	Nice to Have
Urdu	No practical proficiency	Limited working proficiency	Nice to Have
Uzbek	No practical proficiency	Limited working proficiency	Nice to Have
Vietnamese-STD	No practical proficiency	Limited working proficiency	Nice to Have
Tamil	No practical proficiency	Limited working proficiency	Nice to Have